Towards Evaluating World Modeling for Autonomous Navigation in Unstructured and Dynamic Environments

Rolf Lakaemper
Temple University
Philadelphia, PA, USA
lakamper@temple.edu

Raj Madhavan
Institute for Systems Research
University of Maryland, College Park &
Intelligent Systems Division
National Institute of Standards and Technology
Gaithersburg, MD, USA
madhavan@umd.edu/madhavan@nist.gov

ABSTRACT

With funding from the Commerce Department's National Institute of Standards and Technology (NIST) Measurement Science and Engineering Research Grants[1], the authors have recently embarked on a three year project to create and experimentally validate a framework by which automated guided vehicles (AGVs) can automatically generate a sufficiently accurate internal map (world model) of its surroundings. The work presented in this paper discusses challenges involved and reports on a possible extension to a previously-developed mapping technique in evaluating world models of such dynamic and unstructured environments. The paper also reports on the authors' views in bringing together the community to collectively address this problem from end-users', vendors' and developers' points of view.

General Terms

Performance Evaluation, Benchmarking, World Modeling, Manufacturing

Keywords

automated guided vehicles, forklifts, robot mapping and navigation, warehouses, factory floors

1. INTRODUCTION

Having robots sense unstructured environments and automatically generate a sufficiently accurate world model is still an unsolved problem, despite advances in computing power and sensor technologies. The solution requires a framework for generating accurate representations that takes into account the dynamic nature of the operational domain. For example, Automated Guided Vehicles (AGVs) are widely used on factory floors and warehouses to transport goods. Currently they require highly structured environments and reference markers installed throughout the intended domain of operation, which, apart from carrying prohibitively high maintenance and installation costs, are not able to cope with dynamic changes in the environment. This has widespread implications for the applicability of AGVs, and also drastically limits the way modern warehouses and manufacturing floors can be designed. A breakthrough will be achieved if AGVs could cope with unstructured, dynamic environments and adapt to human-centered collaboration. A similar analogy can be extended to various domains.

Measuring the performance of such navigation systems requires scientifically sound and statistically significant metrics, measurement, and evaluation methodologies for quantifying their performance. With funding from the Commerce Department's National Institute of Standards and Technology[2] Measurement Science and Engineering Research Grants, the authors have recently embarked on a three year project to create and experimentally validate a framework by which AGVs can automatically generate a sufficiently accurate internal map (world model) of its surroundings. In addition, the authors are also interested in designing experiments and test methods to enable performance evaluation and benchmarking towards characterizing constituent components of navigation and world modeling systems that provide statistically significant results and quantifiable performance data. The work presented in this paper discusses challenges involved in evaluating world models of such dynamic and unstructured environments.

This paper is structured as follows: Section 2 discusses AGVs and their acceptance within the manufacturing industry. Sections 3 and 4 discuss world modeling and challenges associated with evaluating such models. Section 5 presents our ideas on how to evaluate robot-generated world models. Section 6 concludes the paper by summarizing remaining issues and our continuing work.

[1]The project is jointly carried out by Temple University and the University of Maryland, College Park, under award ARRA-60NANB10D012 to "bolster U.S. scientific and technological infrastructure, increasing our nation's ability to innovate, compete, and solve scientific and technological problems".

[2]Certain commercial equipment, instruments, or materials are identified in this document. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that the products identified are necessarily the best available for the purpose.

2. AGV: INTEREST AND ACCEPTANCE

Robotics and automation holds immense promise as a key transformative technology to positively impact U.S. manufacturing. From traditional and well-established applications in the automotive industry to emerging applications such as material handling, palletizing, and logistics in warehouses, the use of robots can increase productivity whilst ensuring personnel safety. Automated Guided Vehicles (AGVs) represent an integral component of today's manufacturing processes. They are widely used on factory floors for intra-factory transport of goods between conveyors and assembly sections, parts and frame movements, and truck-trailer loading and unloading. To offset prohibitively expensive maintenance and installation costs, and thus expand the AGV's markets and utility beyond what is possible today, it is evident that the dependency on infrastructure is to be minimized (if not eliminated).

A survey on adoption of AGVs conducted in July 2009 by RMT Robotics and Modern Materials Handling Magazine offers some interesting insight into the end-user's opinion about AGVs (see Figures 1, 2). The survey asked end-users about AGV related topics for e.g. desired characteristics of AGVs. While most of the desired characteristics such as durability, low maintenance and adaptability are to be expected, topics related to improved navigation were given a comparably significant attention: open path navigation (78%), obstacle avoidance (91%), no external path guides/sensors (67%). At the same time, the consideration to use AGVs is enormous, see Figure 2. Putting these two results together surely shows that an improvement in navigation capabilities will open up a huge market.

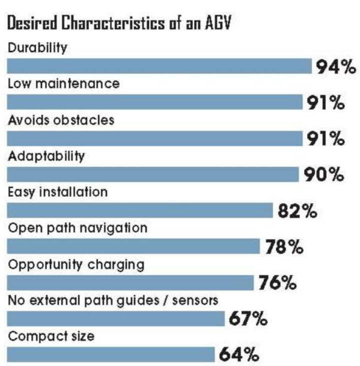

Figure 1: Adapted from an RMT Robotics and Modern Materials Handling Magazine survey. See text for more details.

However, the same survey shows a strong contrast to the interest in AGVs when it comes to familiarity with AGVs. Nearly 50% of the users admitted to be not very/not at all familiar; even 86% have not used or even evaluated the use of AGVs. This discrepancy strongly suggests that the current state of AGVs does not match the needs of end-users. Dependable and robust navigation on factory floors is an important factor to remedy this situation.

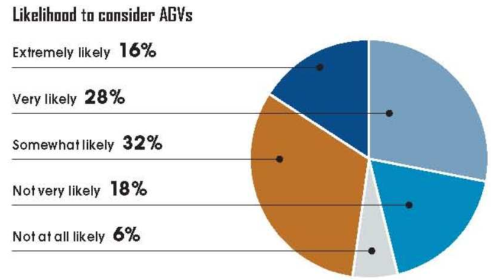

Figure 2: Adapted from an RMT Robotics and Modern Materials Handling Magazine survey. See text for more details.

3. WORLD MODELING

Sensing unstructured environments and automatically generating a sufficiently accurate world model without re- engineering the operating environment is still an unsolved problem despite advances in sensor systems and computing power. To create useful man-machine collaborative systems and to provide continual situational awareness, a framework for generating accurate representations of the operational domain is imperative. Even in the well-established industrial robotics area, it is telling that only five percent of those robots employ sensors as part of their feedback loop [2]. The science of robot vision has gone through significant changes in the past decades. Limitations of purely geometry driven approaches necessitated the need for statistical methods. Although these methods, like Iterative Closest Point (ICP) [1] or Particle Filters [6], generated excellent results in many applications, limitations still exist, e.g. the necessity for a high sampling density, the dependency on odometry, and the utilization of low-level features (pixels/reflection points) only.

These limitations might partially explain the reluctance and hesitation of many AGV vendors to work towards an improvement of the heavily constrained (virtual) track- and marker-based AGV systems. The recently held North American Material Handling Logistics Tradeshow [7], a premier handling event with participation of major AGV and forklift vendors, provided ample testimony to the fact that the majority of AGV developers are reluctant to abandon the track or marker guided paradigm. While only a few companies are beginning to present approaches of track and marker free navigation, most indulge in little engineering improvements like invisible (chemical) tracks, markers with improved visibility, more sophisticated pattern markers, etc. None of the latter approaches are solutions to the previously-mentioned navigation capabilities, demanded by the end-users.

The employment of AGVs is driven by a single parameter: productivity. Only if the application of new technology offers a clear advantage in terms of productivity it will be applied. An interesting example is the application of current AGVs in warehouses: even if (again: mostly track or marker based) solutions of automated vehicles inside warehouses are implemented, there is, in general, no solution for 'the last five meters', the task of loading pallets into trucks.

Compared to the low increase in productivity, the technical challenge and the needed investment into research and development is too high to be appealing as 'low hanging fruit'.

Looking at promising solutions for robot navigation tasks that exist in theoretical publications and, at an academic level, in implementation, robot navigation in constrained unstructured and dynamic environments like factories and warehouses seems within reach. That begs the question: Why aren't we seeing AGVs freely roaming in these environments? We have an interesting discrepancy between academia and industrial implementation (often referred to as the 'real world'):

- Scientists claim to have robust and fast solutions even for seemingly more challenging tasks in robotics Vs Performance is assessed by academia only (peer evaluation of journal papers with stronger emphasis on theory).
- Mapping in (static) indoor environments is often seen as a solved problem by academia Vs Industry acceptable algorithms are seldom implemented.

4. EVALUATION OF WORLD MODELING

Not surprisingly, the development of efficient world modeling schemes has received its due attention from roboticists. A myriad of approaches have been proposed and implemented, some with greater success than others. The capabilities and limitations of these approaches vary significantly depending not only on the operational domain, and onboard sensor suite limitations, but also on the requirements of the end-user: Will a 2D map suffice as an approximation of a 3D environment? Is a metric map really needed or is it sufficient to have a topological representation for the intended tasks or do we need a hybrid metric-topological map [13]? It is thus essential for both developers and consumers of robotic systems to understand the performance characteristics of employed methodologies which will allow them to make an informed decision.

To our knowledge, there is no accepted benchmark for quantitatively evaluating the performance of robotic mapping systems against user-defined requirements. Currently, the evaluation of robotic maps is based on qualitative analysis (i.e. visual inspection). This approach does not allow for better understanding of what errors specific systems are prone to and what systems meet the requirements. It has become common practice in the literature to compare newly developed mapping algorithms with former methods by presenting images of generated maps. This procedure turns out to be suboptimal, particularly when applied to large-scale maps.

Some researchers have recognized the need for quantitative evaluation of mapping and position estimation schemes and are attempting to address it through several programs. For example, there are initiatives to provide collections of standard robotics data sets such as Radish: The Robotics Data Set Repository [9] and RAWSEEDS: Robotics Advancement through Web-publishing of Sensorial and Elaborated Extensive Data Sets [10]) and source codes of various robotics algorithms, such as OpenSLAM [8] and the Mobile Robot Programming Toolkit [12]. While a step in the right direction, they do not address objective performance evaluation and replication of algorithms is not straightforward. NIST's Reference Test Arenas for urban search and rescue robots have been developed to provide the research community with an efficient way to test their algorithms without having to incur the costs associated with maintaining functional robots and traveling to one of the permanent arena sites for validation and practice.

The RoboCup Rescue competitions [11] have proved to be a good forum to evaluate task-based performance of robots. An image similarity metric and a cross entropy metric are outlined in [15] to measure the quality of occupancy grid maps. The metric gives an indication of distortion of the map with respect to a ground truth map in the presence of noise and pose errors. This metric is embedded in the Jacobs Map Analysis Toolkit [14] and has been tested for comparing maps in the RoboCup context. The Jacobs Map Analysis Toolkit, recently extended to evaluate maps using fiducial markers, has major drawbacks since it is purely tailored to perform evaluation of geometric precision, which limits its versatility to be applied to evaluation of maps under different aspects, e.g. the aforementioned end-user requirements (i.e. 2D/3D, topological/geometric mapping).

We provide a different approach to mapping evaluation based on the principles of an algorithm which was originally introduced in [3] as a mapping algorithm: Force Field Simulation (FFS) with Virtual Scans (VS). Although originally created for mapping, a re-interpretation of its core principles leads to an evaluation tool (the 'evaluation mapper') for mapping evaluation with ground truth. The evaluation mapper is tailored for evaluation and not optimized to build maps. Mapping based on FFS-VS is a promising candidate to solve the evaluation task.

5. FFS AS AN EVALUATION TOOL

This section will give a short introduction to FFS&VS with respect to its re-interpretation as a mapping tool. Further details about FFS&VS can be found in [3]. The basic principle of FFS (see Figure 3): driven by attractive forces between features, single scans are iteratively translated/rotated. By laws of physics, such a system converges towards a (local) minimum of its potential.

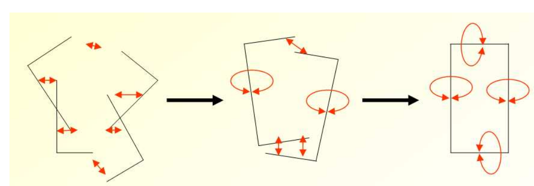

Figure 3: Basic principle of FFS: alignment of single scans is based on attraction-forces, computed from corresponding features.

FFS can be extended using hypotheses of expected features in the environment, so called Virtual Scans, see Figure 4. Currently, these features are mid level geometric objects, like planes, rectangles etc. These objects are detected by a module based on mid level spatial cognition (MLSC), see Figure 4. Real data and Virtual Scans are fed into the iterative Force Field Simulation (FFS) alignment approach. FFS cannot distinguish between real and virtual scans. After each iteration, the resulting global map is re-analyzed to state (new) hypotheses (=create VS).

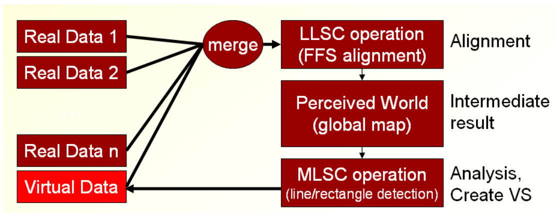

Figure 4: FFS extended by Virtual Scans: While FFS is a Low Level Spatial Cognition (LLSC) based module, the map analysis is based on Mid Level Spatial Cognition (MLSC) principles. Virtual scans augment the data gained from sensors.

Interestingly, FFS&VS can align a set of local maps (e.g. single scans) towards a 'ground truth'. With a little re-interpretation, this can be used for map-evaluation, making the tool an evaluation-mapper:

- Assuming a ground truth map G, and the result R of a mapping algorithm under evaluation, proceed as follows:
- Change the role of G and R: R becomes ground truth, and is used as Virtual Scan.
- Manually split G into logic (intuitive, motivation follows) parts.
- Use FFS to align G towards R.
- Evaluate the alignment effort G towards R, leading to a quality measure of R.

The split of the ground truth map and its alignment to the evaluated map is a major difference between the proposed approach and the static approach used by the Jacobs Toolkit. A more general view on map evaluation highlights the versatility gained by this step.

5.1 Grid-based Evaluation

In grid-based evaluation, the map to be evaluated (target map) and the ground truth map are both embedded into a common grid. The grid cells are labeled using properties like 'object', 'empty space' or 'hidden'. Tools using this approach, e.g. the Jacobs Toolkit, measure the local geometric accuracy of the map. Since only low level features (object/empty space) are incorporated, the target map must be close to the ground truth map: it is assumed that low level correspondences imply higher level correspondences. Larger errors in the global appearance of maps cannot be quantified; globally erroneous maps are classified as 'wrong' - even if they are locally correct.

Figure 5 illustrates a case where global geometric correctness might be of minor interest compared to only locally geometric, yet globally topological accuracy. This is the case for optimal path-planning algorithms for AGVs. Figure 5, left, shows the ground truth map. Figure 5, center, illustrates a mapping result with high global geometric correctness, although the bottom part is wrong in details. Figure 5, right, is an example for a map with a high global geometric error. However, all details (obstacles, target position in bottom section) are mapped correctly, the map is also topologically correct (two rooms are connected by hallways). A grid-based approach will prefer the center map to the right one. If the map is intended for AGV navigation, the right map is of higher quality: it shows correctly that the target (red dot) is reachable from the current position (black dot) using the right hallway. The center map misleads the AGV to take the left hallway, a probably expensive mistake. Grid-based approaches such as the Jacobs Toolkit aim to measure the global topographic quality of a robot map; they cannot quantify the topological qualities of a map.

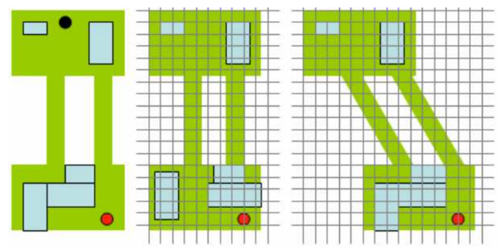

Figure 5: Grid-based evaluation. Left: Ground truth map. Two rooms (green, top and bottom) with obstacles (gray) are connected by two hallways (green, center). The target (red dot) can only be reached from the current position (black dot) using the right hallway. Center: Mapping example with high global and low local correctness. Right: Mapping example with low global and high local correctness. A grid-based map evaluation will prefer the center map, although for AGV navigation the right map is of higher utility.

5.2 Pose-based Evaluation

A different approach to mapping evaluation is pose-based map quality estimation [4]. Pose-based fitness exploits the fact that precise robot localization is dual to robot mapping: if the robot pose is precisely known in the ground truth map, the scans can be registered based on the pose estimates. Since robot pose measurements are imprecise, the scan data itself has to be taken into account to register the scans in a common coordinate system. Successful registration of scans adjusts the robot poses defined by the target map into the ground truth coordinate system.

Evaluation based on pose information compares the adjustment of robot poses; the sum of all pose errors yields the overall error. The main advantage of pose-based evaluation is its applicability in higher dimensions (for e.g. in 6D-Simultaneous Localization and Mapping schemes). The number of poses to be evaluated is dimensionality independent, whereas the memory consumption of a grid-based evaluation approach increases for 3D applications to a prohibitive cubic behavior. Hence there is high interest in gaining knowledge about pose-based evaluation. The global topological correctness is captured by the fact that only a few rotations are needed to achieve the optimal result. Pose-based approaches have certain drawbacks, mostly related to the fact that rotational and translational errors have significantly different perceptual effects, but are handled alike in the pose error computation.

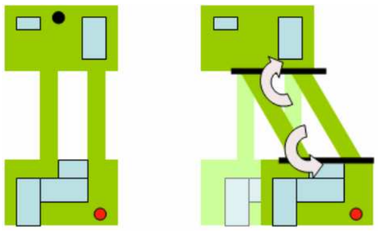

Figure 6: Pose-based evaluation. The target map (right) is transformed to match the ground truth (left). The transformation parameters (here: rotation, arrows) are used to quantify the map quality. The topological correctness of the target map is reflected by the fact that only two rotations are needed to achieve the optimal map.

5.3 Hybrid Evaluation

FFS&VS, if used as evaluation-mapper as described above, can be utilized for combined evaluation with respect to topographical and topological map properties, which results in a hybrid pose/grid-based evaluation methodology. Emerging from the proposed framework, Virtual Scan assisted Force Field Simulation, it is designed to eliminate the drawbacks of pure pose- or grid-based evaluation schemes while exploiting their advantages. Using the target map as a fixed virtual scan with high confidence weight, it aligns the single scans of the decomposed ground truth map to the target map, see Figure 7.

Observe that in this approach we transform the ground truth map, not the target map. There are two reasons for such an approach: first, it makes the evaluation independent of the target map's data format. Since the target map is not transformed, it can be given in any format (e.g. geotiff). Second, and perhaps more importantly, the part-decomposition of the ground truth map can reflect the task specific requirements of the world modeling framework. For example, the ground truth map of Figure 7(a) is decomposed into top room, hallways and bottom room. These three parts are required to be mapped with high geometric accuracy. The hybrid approach quantifies the map quality using pose-based parameters and grid-based parameters from additional evaluations on the single parts. Their relative pose, defines the global appearance of the map, which is captured by the transformation parameters. The importance weight of the transformation parameters can be individually determined.

6. CONCLUSIONS AND FUTURE WORK

This paper presented some preliminary thoughts on how to evaluate robot-generated world modeling schemes for unstructured and dynamic environments. The use of world modeling for achieving autonomous navigation of AGVs and forklifts on manufacturing floors and warehouses was elaborated. Many open questions remain: What is the industry's view on the discrepancy between the quality assessment of algorithms of academia and industry? What are industry's requirements? What are typical cases of environments where AGVs are needed? Which of the different mapping approaches could be upgraded to live up to industry standards?

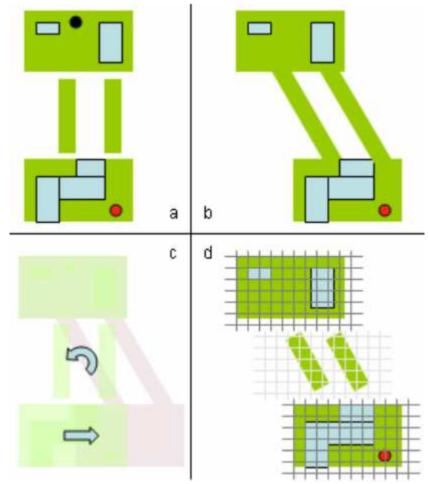

Figure 7: Hybrid evaluation methodology. (a) Decomposed ground truth map (3 parts). (b) Target map. (c) Mapping transformation of (a) to (b). (d) Grid-based evaluation on transformed parts. The final score is computed using task adjusted weights for transformation parameters and grid evaluation results.

The paper also emphasized the need for objective evaluation via development of test methods to quantify the quality of world models. Our ongoing work is focused on developing benchmarking schemes and how these can be channeled towards facilitating the development of standards for the AGV industry [5].

In an effort to bring together the academic and research communities together, the authors have organized two workshops, held in early September 2010 at Temple University and at the 2010 Performance Metrics for Intelligent Systems (PerMIS) Workshop where these issues were discussed. The discussion focused on the challenges in achieving these goals. Technical presentations and open discussions centered on how to create and experimentally validate world modeling frameworks for unstructured environments amidst dynamic objects.

7. REFERENCES

[1] P. Besl and N. McKay. A method for Registration of 3D Shapes. *IEEE PAMI*, 14, 1992.

[2] CCC/CRA. Roadmapping for robotics. http://www.us-robotics.us.

[3] R. Lakaemper. Improving sparse laser scan alignment with virtual scans. In *International Conference on Intelligent Robots and Systems (IROS08)*, Nice, France, September 2008. IEEE.

[4] R. Lakaemper, A. Nuechter, and N. Adluru. Performance of 6d lum and ffs slam. In *Performance Metrics for Intelligent Systems (PerMIS) Workshop*, Gaithersburg, MD, 2008.
[5] R. Madhavan, R. Lakaemper, and T. Kalmar-Nagy. Benchmarking and standardization of intelligent robotic systems. In *14th International Conference on Advanced Robotics (ICAR 2009)*, Munich, Germany, June 2009.
[6] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit. Fastslam 2.0: An improved particle filtering algorithm for simultaneous localization and mapping that provably converges. *Proc. of the Int. Conf. on Artificial Intelligence (IJCAI)*, 2003.
[7] NA. North American Material Handling Logistics Tradeshow, Cleveland, Ohio, April 2010.
[8] OpenSLAM. http://www.openslam.org.
[9] Radish. The robotics data set repository. http://radish.sourceforge.net.
[10] RAWSEEDS. Robotics advancement through web-publishing of sensorial and elaborated extensive data sets. http://www.rawseeds.org.
[11] Robocuprescue. http://www.robocuprescue.org.
[12] M. Robots. The mobile robot programming toolkit. http://babel.isa. uma.es/mrpt/.
[13] S. Thrun. Learning metric-topological maps for indoor mobile robot navigation. *Artificial Intelligence*, 1998.
[14] J. University. Jacobs map analysis toolkit. http://robotics.iu-bremen.de/datasets/mapevaluation.
[15] I. Varsadan, A. Birk, and M. Pfingsthorn. Determining map quality through an image similarity metric. In *Lecture Notes in Artificial Intelligence (LNAI): RoboCup 2008,Robot WorldCup XII*, 2008.